***Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?***

Looking at the bar graph created from the pivot table comparing parent category to outcome, we see that the theater category had the highest amount of crowdfunds created (344 out of 1,000) with the most failed (132 out of 364) and the most successful (187 out of 565). At first glance, it looks as if theater performed the best, but further testing would be required to evaluate if this is a significant difference compared to the other categories. We also see that journalism had the lowest count of campaigns created (with a total of 4), had a success rate of 100%, which is something that no other category accomplished. When looking at the bar graph that compares sub-categories to outcome, we see that plays (which makes up the entirety of the parent category of theater) had the highest amount of crowdfunds created, most failed, and most successful. This does not tell us much about theaters/plays as they have high numbers of successes and high numbers of failures. All in all, we see that almost all categories had more successes than failures, with a total of 565 successes out of 1,000. This applies all year round as we see when we compare outcome to dates created by month. So, not much can be said for what time of year is better for crowdfunding campaigns. This also applies for most goal ranges, with the exception of campaigns in the $10,000-$14,999 range and the more-than-$50,000 range. Because the ranges are on different spectrums (one on the lower end of the range and one on the high end), it’s difficult to tell if goal range plays a significant role in campaign outcome. The argument can be made that perhaps, due to a high number of reported successes, crowdfunding may be a great alternative to gathering funds for any given project. This claim is substantial and would require further analysis and testing.

***What are some limitations of this dataset?***

This is a reasonably sized data set comprised of 1,000 projects that have been separated into many small categories, that the same sizes of most categories is simply too small to hold any statistical significance. This makes it difficult to draw any strong conclusions from the collected data. It is also unknown how promotion played a role in collecting funds. There is no reported way to measure how these campaigns were shared across different types of platforms, advertised, and marketed. If a campaign is shared more frequently, it may be likely that more people will contribute, and vice versa. The way in which it is marketed can also play a role in how likely a person may be to donate.

***What are some other possible tables and/or graphs that we could create, and what additional value would they provide?***

One important factor that plays a role in the success of a campaign is the amount of time the project was live for. It can be hypothesized that the longer a project is live, the more donations it will receive, given it is well promoted. Just as if a project is live for a short period of time, many may not reach it in time and miss out on potential donations. A line graph comparing amount of time live to outcome can show whether this idea applies. Another interesting analysis to make would be to see the average donation made by backers compared to categories to evaluate how much people are willing to fund different fields. This can be seen through a bar graph comparing average donation amount to category. Additionally, a bar graph comparing the amount of backers to the outcome of a project can give us an idea of how many backers it takes to reach a goal on average. This can also be filtered by range.

STATISTICAL ANALYSIS

***Use your data to determine whether the mean or the median better summarizes the data.***

The median would be a better and more representative number of the sample for both successful and failed backers. When looking at the successful backers, we see a range of 16 to 7295 with a standard deviation of 1266. The mean of this group was 851, heavily skewed to the higher end of the data. The median was 201, giving us a better idea of what the sample represents, given its large sample size. As for the unsuccessful backers, we see a range of 0 to 6080, with a standard deviation of 960. The mean was 568 and the median was 115. The same concept applies, which is why the median of 115 better represents the sample.

***Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?***

At first, it looks that there is not a great difference in variability between successful and unsuccessful campaigns. Both are large data sets, with large ranges between min and max amounts, high standard deviations, and high variance. It seems that there is greater variability in the successful campaign as there were more successful campaigns than there were failed, so it makes sense that we would see bigger differences. Further testing would be needed to determine statistical significance, but both campaigns vary greatly.